資料１－１

初等中等教育段階における生成AIの利活用に関する検討会議

生成AIの現在地と未来
及び
教育への影響

株式会社GenesisAI 代表取締役社長/CEO
博士（工学）
今井翔太

自己紹介

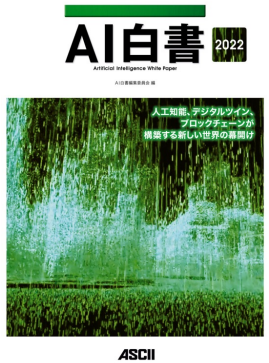
今井 翔太（Shota Imai）

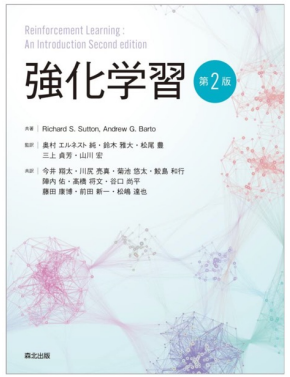
1994年，石川県金沢市生まれ

2024年3月：東京大学大学院 工学系研究科技術経営戦略学専攻
松尾研究室にて博士（工学）を取得

2024年7月：株式会社GenesisAI創業 代表取締役社長/CEO

- 研究分野：強化学習，マルチエージェント強化学習，LLM
- 著書
  - **生成AIで世界はこう変わる**
  - G検定公式テキスト 第3版
  - AI白書2022
  - Sutton著『Reinforcement Learning』翻訳





Twitter：
@ImAI_Eruel
公式サイト：
http://shota-imai.com

本発表について

- 本資料は，企業の代表としてものではなく，研究者の視点からの内容になります
- 本資料で参照している研究は2024年8月1日時点のものです
- 本会議には，本資料で参照する研究を実施する生成AI研究の最先端の事業者，研究者が複数出席していることを鑑み，公平・フラットな視点での解説になります

現在の生成AIの到達点

■ **言語**
- 30近い言語を使用可能（GPT-4~）
- 医師国家試験に合格可能（GPT-4，2023年時点）
- 司法試験に合格可能（GPT-4，2023年時点）
- あらゆる分野に関して大学生レベルの知識（GPT-4o，Gemini 1.5，Claude3.5）注1
- いくつかの学問分野で大学院生レベルの推論能力（GPT-4o，Gemini 1.5，Claude3.5）注2
- 1000万文字（※正確には文字ではなくトークン）以上を瞬時に読み取り，要約や質問回答（Gemini 1.5）
- 昨年のGPT-4レベルのモデルが個人のPC上で動作可能（オープンモデルの一部，Gemma2，Llama3.xなど）

■ **動画像**
- 文字レベルの細かい描写（拡散モデル系，GPT-4o）
- 2つ以上の動画像の自然な結合（拡散モデル系，Sora，GPT-4o）
- 分単位で，文章から人間が現実と見分けがつかない動画を生成可能（Sora， Veoなど）
- 上記について，静止画を意図通りに動かすことも可能
- 3Dモデルの生成（GPT-4o，Stable Video 3Dなど）

■ **音声**
- 数秒程度の声のサンプルから当人と同じ声で自由なセリフの発声（GPT-4o，VALL-E-Xなど）
- 効果音・自然音の生成（AudioCraft，Stable Audioなど）
- 長時間の音楽の生成（AudioCraft，Stable Audioなど）
- 歌詞入力のみから歌唱&伴奏の生成（Suno AI）
- 声による自然な感情表現（GPT-4o）

■ **マルチモーダル**
- ラフなメモ書きからウェブサイトのコード生成（2024年，GPT-4o，Gemini 1.5，Claude3.5）
- 10時間以上の動画を瞬時に処理し要約や質問解答（Gemini 1.5）
- 107時間以上の音声を瞬時に認識し要約や質問解答，文字起こし（Gemini 1.5）

■ **その他（生成AI + αの手法）**
- 数学オリンピックで銀メダルレベル（AlphaGeometry + AlphaProof，LLM+探索手法）
- プロンプトのみから，ツールを駆使し複雑なソフトウェアの設計（Devin，AIエージェント系）

	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro	Llama-400b (early snapshot)
Graduate level reasoning *GPQA, Diamond*	59.4%* 0-shot CoT	50.4% 0-shot CoT	53.6% 0-shot CoT	–	–
Undergraduate level knowledge *MMLU*	88.7%** 5-shot 88.3% 0-shot CoT	86.8% 5-shot 85.7% 0-shot CoT	– 88.7% 0-shot CoT	85.9% 5-shot –	86.1% 5-shot –
Code *HumanEval*	92.0% 0-shot	84.9% 0-shot	90.2% 0-shot	84.1% 0-shot	84.1% 0-shot
Multilingual math *MGSM*	91.6% 0-shot CoT	90.7% 0-shot CoT	90.5% 0-shot CoT	87.5% 8-shot	–
Reasoning over text *DROP, F1 score*	87.1 3-shot	83.1 3-shot	83.4 3-shot	74.9 Variable shots	83.5 3-shot Pre-trained model
Mixed evaluations *BIG-Bench-Hard*	93.1% 3-shot CoT	86.8% 3-shot CoT	–	89.2% 3-shot CoT	85.3% 3-shot CoT Pre-trained model
Math problem-solving *MATH*	71.1% 0-shot CoT	60.1% 0-shot CoT	76.6% 0-shot CoT	67.7% 4-shot	57.8% 4-shot CoT
Grade school math *GSM8K*	96.4% 0-shot CoT	95.0% 0-shot CoT	–	90.8% 11-shot	94.1% 8-shot CoT

* Claude 3.5 Sonnet scores 67.2% on 5-shot CoT GPQA with maj@32
** Claude 3.5 Sonnet scores 90.4% on MMLU with 5-shot CoT prompting

出典：Anthropic“Claude 3.5 Sonnet（https://www.anthropic.com/news/claude-3-5-sonnet）”．最先端のClaude，GPT，Geminiの最新モデルは，MMLUやGPSQなどの言語性能のベンチマークで専門家の人間レベルの領域で性能を争う．なお，この図（6月後半）における比較は古く，現在(8月)は入れ替わりが発生していることに注意（後述のLMSYSのリーダーボードを参照されたい）

- 注1：MMLUで80%代後半のスコア．MMLUはLLMの評価で最も使用されるベンチマークであり，人分，社会科学，自然科学などの大学学部レベルの選択問題からなる．人間の専門家で89.8%．
- 注2：GPQAで50%代後半のスコア．生物学，物理学，化学の各分野の専門家によって作成された大学院レベルの難問セット．特定領域の博士号レベルの人がGPQA内のその領域の問題で65%の精度．専門分野以外ではインターネットへの無制限アクセスを許可した場合でも34%

直近の各分野における生成AIの動向

【言語，マルチモーダル】

～2024年：言語，マルチモーダル性能共にGPT-4の天下．モデルサイズ大規模化の傾向．

- 2024年5月14日：GPT-4o発表（OpenAI）
  →極めて高速かつ，言語，動画，画像，音声全ての「入出力」が可能になった初のモデル．感情表現まで可能
- 2024年5月15日：Gemini 1.5 Pro発表（Google）
  →最大で1000万トークン処理可能，動画を10時間以上，音声も100時間以上処理可能な初のモデル
- 2024年5月18日：小型LLM「Phi」をWindows PCに搭載すると発表（Microsoft）
  →デフォルトでPCに生成AIが搭載される時代に
- 2024年6月10日：iPhone上で動作する生成AIサービスApple Intelligenceを発表（Apple）
  →スマホ上で生成AIサービスが提供される時代に
- 2024年6月21日：Claude 3.5シリーズ発表（Anthropic）
  →GPT-4oを上回る性能を持つ初のモデル
- 2024年7月1日：Gemma2公開（OpenAI）
  →昨年GPT-4レベルの性能でローカルPCで動かせるモデル
- 2024年7月23日：Llama3.1公開（Meta）
  →オープンモデルでGPT-4oを上回る初のモデル
- 2024年8月1日：Gemini 1.5 Proの最新版がGPT-4oやClaude3.5を上回り，全てのLLMの中でトップの性能に（Google）

【その他（生成AI＋α）】

~2024年：生成AIがツールを使いながら自律的に目標を達成したり推論能力を上げる手法の総称としてAIエージェント（あるいはLLMエージェント）という概念が成立．マルチエージェントの手法も出現．

- Devin（Cognition）
  →単なるコード生成だけでなく，計画，ドキュメントの参照，自動的なエラー解決，体系的なソフトウェアエンジニアリングを自動的にこなせる初のAIエージェント
- AlphaGeometry＋AlphaProof（Google）
  →数学オリンピックで銀メダルレベルになった初のAI
- Q*，GPT-5（OpenAI）
  →OpenAIが開発を公言（あるいは噂されている）しているAI．生成AI+αで推論能力を挙げたという説も

【動画像】

～2024年：オープンなStable Diffusion，高品質のMidjourney，クリーンかつコントローラブルなAdobe Fireflyなどが台頭．文字を正確に描写可能なDALL・E3など弱点の克服も．

2024年2月：Sora発表
→初の長時間生成&実用レベルの動画生成AI．拡散モデル+Tranformerの組み合わせでLLM的な大規模学習が動画でも有効なことを示す

- 2024年5月：GPT-4o発表（OpenAI）
  →画像生成や画像同士の結合，3Dモデル生成も可能
- 2024年5月：Veo発表（Google）
  →Soraに匹敵する初の動画生成AI
- 2024年6月：Kling公開（中国 快手）
- 2024年7月：DreamMachine公開（Luma）
- 2024年7月：Gen-3公開（Runway）
  →ビッグテック以外でSoraレベルの動画生成AIの開発が進み一般利用可能に
- 2024年7月：StableDiffusionレベルのAIが超低コスト（数十万円程度）学習可能な手法発表（Sony AI，カリフォルニア大）

【音声】

~2024年：効果音や音楽単体の生成（AudioCraft），人の声の再現（RVC, so-vits-svc等），歌唱+伴奏の生成（Suno AI）等，着実に進歩．声の生成に関して企業はやや慎重な姿勢．

- 2024年4月：Voice Engine公開（OpenAI）
  → テキストと15秒の音声サンプルから当人の声で好きな読み上げ
- 2024年5月：GPT-4o発表（OpenAI）
  →声で感情表現，効果音なども生成可能
- 2024年6月：Stable Audio Open（Stability AI）
  →Meta社のAudioCraftに続き，実用的が音楽生成のオープンモデル
- 2024年7月：SunoAIで伴奏と歌唱の分離可に（Suno）
- 2024年7月31：GPT-4oボイスモード解放（OpenAI）

直近の生成AIに関する研究上の重要知見と発展傾向

- **オープンモデルの進化→生成AIの民主化**
  - Gemma2やLlama3.1のように，誰でもインストール&カスタマイズ可能なオープンモデルの性能がほぼ最先端のモデル（GPT-4等）に追いついた
- **言語モデルの高速・小型・効率化→エッジデバイスへの生成AI搭載搭載**
  - Gemma2（20億～270億パラメータ，Google），Phi-3（38億~140億パラメータ，Microsoft）等のPCやスマホ上で動かせるモデルは1年前の数百億～1兆パラメータのモデルに匹敵する性能に
  - 大型モデルの出力をまねる蒸留の手法や，スケーリング則やチンチラ則等の説で明らかになっていたモデルサイズ-データ比率を超えた学習により，小さいモデルに知識を詰め込めるようになった
  - ニューラルネットワークのパラメータの量子化技術（例：BitNetはパラメータを[1，0，-1]の3値のみで表現）により，既存モデルを低メモリ，低遅延，低消費電力で利用可能に
  - GPT-4oやGemini Pro Flashのように，（詳細な技術は公開されていないが）Transformerのアーキテクチャレベルでの工夫により極めて高速な推論が可能に
- **非学習手法による生成AIの高性能化→ビッグテック以外にも高性能生成AI開発の道**
  - 進化計算により複数の生成AIの層を組み合わせていわゆる「いいとこ取り」のモデルを作る手法（進化的モデルマージ）や重みを足すだけでLLMに対話能力を与える手法（ChatVector）などが確立
- **マルチモーダルネイティブな生成AI→あらゆる種類の情報をそのまま処理できるように**
  - 従来は，言語のみで大規模学習をおこなったあとにマルチモーダル機能を追加したり，言語以外の情報をテキストトークンに変換してから処理していたため，言語以外の性能低下や処理遅延などが存在したが，現在は最初からマルチモーダルなデータで学習
- **高性能な生成AIの乱立→生成AIの再現性**
  - Soraに続いてすぐGoogleのVeoや，ビッグテック以外から複数の動画生成AIが出たこと，上記のオープンモデルの進化のように，前提条件さえ揃えば誰でも高性能な生成AIを作れることが明らかになった．OpenAI一強時代の終焉

Rank* (UB)	Model	Arena Score
1	Gemini-1.5-Pro-Exp-0801	1300
2	GPT-4o-2024-05-13	1286
2	GPT-4o-mini-2024-07-18	1280
4	Claude 3.5 Sonnet	1271
4	Gemini-Advanced-0514	1266
4	Meta-Llama-3.1-405b-Instruct	
5	Gemini-1.5-Pro-API-0514	1261
6	Gemini-1.5-Pro-API-0409-Preview	1257
6	GPT-4-Turbo-2024-04-09	1257
10	GPT-4-1106-preview	1251
10	Claude 3 Opus	1248

誰でもインストールしてカスタマイズできるオープンモデルが，GPT-4やGeminiなどの最先端モデルに匹敵

26	Gemma-2-9B-it	1187
26		1187
26		1186
26		
26	Qwen-Max-0428	
30	DeepSeek-Coder-V2-Instruct	
30	Claude 3 Haiku	1178
34	Meta-Llama-3.1-8b-Instruct	1166
34	Reka-Flash-Preview-20240611	1165
34	GPT-4-0613	

現在のローカルPCで動かせる生成AIの最先端（たった90億パラメータ）

パラメータ数で，約200倍の差

昔のGPT-4の性能（1兆8000億パラメータとされる）

出典：ユーザー投票によりLLMをランキングするサイト
LMSYS Chatbot Arena Leaderboard（https://chat.lmsys.org/）

現時点の生成AI（あるいは一般的な意味でのAI）の弱点

■ ハルシネーション（誤った出力）をしてしまうこと

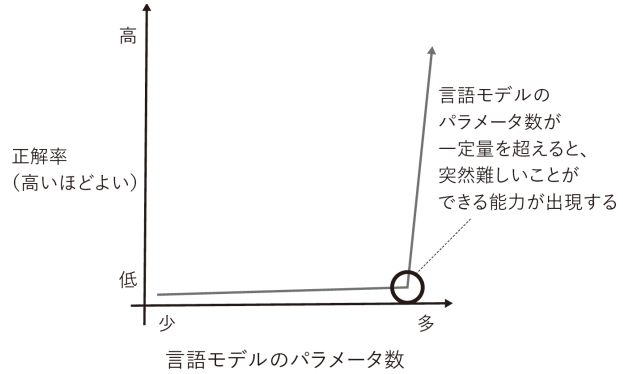
→現時点では研究上，完全に克服する方法は見つかっていない．あるいは言語モデルの仕組み上は不可能？

■ 論理推論，計算に弱い

→大規模化による能力創発[注1]の影響と，プロンプトエンジニアリング，外部ツールの利用により，飛躍的に性能上昇中

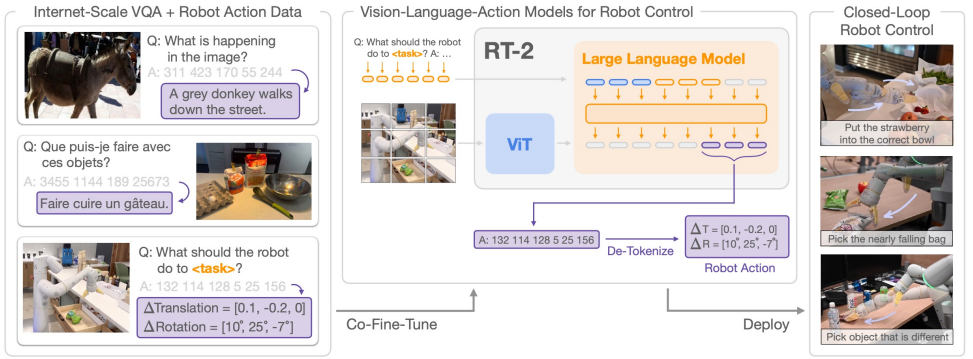
■ 複雑な運動はできない

→スキップや箸を使うなどの基本的な動作も難しい（作り込めば可能だが，汎用性に欠ける）．現在データが少ないだけで大規模学習で解決できる可能性[注2]，そもそも現在のニューラルネットワークの構造では不可能など，様々な説が存在



言語モデルのパラメータ数

注1：言語モデルのパラメータ数が一定量を超えると，非連続的に突然新たな能力が開花する現象．図は元論文”Emergent Abilities of Large Language Models”を参考に作成した今井翔太著『生成AIで世界はこう変わる』から引用



注2：Googleによる，生成AIのように大規模なデータを使うことでロボット制御学習を行う基盤モデルRT（Robotics Transformer）シリーズや，ロボット用の大規模学習データを整備するOpen X-Embodimentプロジェクト（東大も参加）が進行中．図はRT-2の論文” RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control”より

数年後（~5年），生成AI（あるいは一般的な意味でのAI）はどうなるか

※：5年を超える科学の発展，特にAIという極めて進捗が早い分野で発展を予測することは困難で，それを予想として明言するのは研究者として本来あまり誠実ではない（例えば，2020年に「3年後のAIの発展の予測をしろ」と言われて，現在の生成AIブームを当てられた人はほぼ皆無だっただろう）．以下は可能な限り現在の研究から推測した今井個人の予想である

- ホワイトカラー的な能力に関しては，ほぼ全ての能力で人間を超える．LLMの知識が全分野において人間の博士号取得者レベルになる

→投入する学習資源（データ，モデルサイズ，計算量）を増やすことで性能上昇が約束されるスケーリング則[注1]の存在と，ツールを組み合わせるAIエージェントの手法による

- 肉体労働に関しても，相当部分の作業ができるようになる

→2024年時点で，研究者たちがロボティクス分野での大規模学習の準備[注1]や，ニューラルネットワークのアーキテクチャの改善に取り組み始めている．5年後には一定の成果

- 自動研究の発展

→いくつかの機関やプロジェクトで自動研究の取り組みが始まる．特にデジタル空間のみで解決する分野（コンピュータ科学や数学など）に関しては，相当部分が自動化可能に

注1：ほぼLLMの性能向上の文脈で参照されるスケーリング則だが，厳密にはテキスト生成だけでなく，その他の動画像などの生成，マルチモーダル性能にもほぼ同様の法則があることが明らかになっている．Henighanらによる” Scaling Laws for Autoregressive Generative Modeling”など

生成AIの教育へ使い方，影響

■ 実用的な視点での利用法

- 板書や手書きのノートの写真等の非構造情報をマルチモーダル機能で構造化されたテキストに変換
- GPT-4o等の音声会話機能で英語のスピーキング練習を行う
  （※2024年8月1日時点で，GPT-4oのボイスモードが英会話練習で実用的）
- 音声，動画認識機能で，講義内容を全て文字起こし，あるいは要約する
  （Gemini 1.5 Proのマルチモーダル機能であれば既に可能）
- セキュリティ・データの慎重な扱いが求められる場面で，教員や生徒のPC，iPadなどのローカル端末で動く小型のLLMを利用する
- 事実関係があまり問題にならない分野で，視点を広げたり，自分専用の練習問題を作る
- 各教科の学習段階の最初と終了段階でそれぞれ生成AIを使わせ，生徒の習熟度が上がった終了段階の方が生成AIを使った問題解決能力が上昇していることを実感させる

■ 大局的な視点での生成AIの影響

- 従来であれば教員との時間的・物理的制約，あるいは感情的な配慮により教育機会が不十分だった分野に関して，生成AIの利用により習熟度が向上
- 生成AIを補助的に使うことにより自力で解決できる課題が増えることで自信がつき，学校で教えられた分野の熟達や，未知の分野へのチャレンジ（プログラミングによるアプリ作成等）につながる
- 「既に存在する問題解決能力（テストなど）」はAIが人間を超えるようになり，子供たちが無力感を感じてしまう可能性
- 意図せずフェイク情報を生成することで誤った知識を持つ可能性

AI研究者としての個人的な意見
生成AI時代に子供たちに対する教育はどうすべきか？

【考えられる生成AI時代の未来】

- 今ある問題を解く「問題解決能力」に関して，人は現在/未来のAIには勝てない．人間は，人間世界における「解決すべき問題を見つける」ことが役割になる
- ほとんどの人間の能力差は高性能なAIを誰もが使うことによってあまり意味をなさなくなる可能性．能力の大小というよりは，本質的な人間力（信用，人格，コミュニケーション）によって，評価される機会が増える
  - 一方で，真の意味での専門家・トップ層はAIと比べた能力の大小に関係なく評価され（現在でもスポーツ，将棋などの競技，イラストレータ等）る．人間社会で解決すべき課題を見つけるのも，AIの出力の価値判断を行うものも，AIに置き換えられない仕事
- しかし，AIが人間の使うツールである限り，AIの性能は使う人間の能力にも依存する
  - 特に，現在の生成AIは典型的な「使う人間によって性能が大きく変わるAI」である．従来の画像認識AIやゲームAIなどは小学生が使ってもAI研究者が使っても性能は変わらない

【意見】

- 「AIが仕事を奪うのではなく，AIを使いこなす者によって仕事が奪われる．よって教育段階では，従来教科よりAIの使い方を積極的に学ぶべき」との言説があるが，個人的には賛同できない．むしろAIを使いこなすには人間自信の素の能力が必要であり，義務教育レベルの知識は確実に身につけることを前提にAIに関して最低限の教養があればよい
- 専門知，あるいは人間世界の解決すべき問題発見能力を身につける準備として義務教育を機能させる（義務教育レベルの知識がなければ，専門知は身につかない）
- 人間の不完全さ，人格，コミュニケーションこそが人間の本質であることを強調する